NUI Galway
OÉ Gaillimh

Spring Examinations 2011/ 2012

Exam Code(s)	3IF
Exam(s)	3rd year Information Technology
Module Code(s)	CT332
Module(s)	Database Systems II
Discipline(s)	Information Technology
Internal Examiner(s)	Dr. Michael Madden Dr. Colm O'Riordan*
External Examiner(s)	Prof. Michael O'Boyle
No. of Pages	5
Duration	3 hours
Instructions:	Answer any three questions

PTO

Q.1.

i) Explain what is meant by the term *normalisation*.

Given the following relationship, R, and set of functional dependencies, **F**:

R = {A,B,C,D,E,F,G,H,I}

F = { {C,D} -> {A},
{G} -> {E},
{C,D,E} -> {G,B,H},
{B} -> {F},
{H} -> {I} }

decompose R into relations that satisfy Boyce-Codd normal form. *(12)*

ii) With reference to the set of functional dependencies given in part i),

a) Show, using Armstrong's axioms, that the functional dependency GCD-> F is supported (or not) in **F.**

b) Calculate the closure of the set {E}. *(8)*

iii) Describe a procedure you might adopt to map relationships (1:1, 1:N, N:M) from an ER diagram to a relational database schema.
How would the process differ if you were mapping to an OO database schema.
Illustrate your answer with suitable examples. *(13)*

PTO

Q.2.

i) Explain briefly the importance of concurrency control in multi-user databases. Outline the problems that may arise if concurrency control is not enforced. *(6)*

ii) Time-stamping and two-phase locking are two approaches guaranteeing conflict-serializability. Outline *either* approach and present pseudo-code for the primitives used.

Show how the following schedule would proceed under *either* protocol (timestamping or two phase locking).

If adopting two phase-locking, use shared and exclusive locks.

Ta	Tb	Tc
read_item(x)		
read_item(y)		
	read_item(y)	
		read_item(z)
		read_item(x)
write_item(y)		
	write_item(y)	
		write_item(z)

(17)

iii) Explain what is meant by the commit point of a transaction and explain its importance in recovery mechanisms. Explain the differences in recovery processes for systems operating under the deferred update protocol and the immediate update protocol. *(10)*

Q.3.

i) Given the following schema for a database to store information regarding suppliers supplying parts to different jobs.

SUPPLIER: sno, sup_fname, sup_sname, status, location
PART: pno, pname, colour, part_desc
JOB: jno, jname, job_desc, job_loc
SHIPMENT: sno, jnumber, partno, quantity, price

The **SUPPLIER** relation is used to maintain information on all suppliers. The **PART** relation stores information relating to various parts that may be supplied by suppliers. The **JOB** relation is used to store information on jobs to which suppliers may supply parts. The **SHIPMENT** relation captures information pertaining to the ternary relationship between **SUPPLIER, PART** and **JOB.**

Develop SQL code to satisfy the following information need:

List all jobs in London that use a green part supplied (with a quantity greater than 12) by a supplier based in Paris. *(3)*

Describe the process of heuristic optimisation and show how a canonical tree representing the SQL query developed can be transformed to a tree representing an equivalent more efficient query tree. *(14)*

ii) Discuss the structure of a B tree and describe the algorithm for insertion of values into a B tree. Given the values: 11, 15, 13, 35, 44, 22, 32, show how a B-tree of order 3 would develop as the numbers are inserted. *(10)*

iii) Outline briefly, the advantages of adopting a B+tree over a B-tree. *(6)*

Q.4.

i) Choosing appropriate facts, illustrate how the standard relational algebra operators (project, union, join) can be implemented in Datalog. *(10)*

ii) Propose a suitably efficient algorithm for implementing the join operator. How may this be extended to a parallel architecture? *(12)*

iii) In distributed databases, the semi-join operator is often used to implement a join between two relations stored at separate sites. Explain, with the use of an example, how the semi-join operator functions. (*11)*